

Drift and mutation with a finite number of allelic states

(unequal mutation rates/genes alike/transitional and equilibrium values/differentiation of populations)

C. CLARK COCKERHAM

Department of Statistics, North Carolina State University, Raleigh, NC 27650

Contributed by C. Clark Cockerham, September 26, 1983

ABSTRACT The frequencies with which alleles are alike within (Q) and between (q) populations are formulated for monoecious populations undergoing drift and mutation with unequal mutation rates among alleles for a finite number (k) of allelic states. The effective number of alleles and an application to Nei's measure of genetic distance [Nei, M. (1972) *Am. Nat.* 106, 283-292] are also considered for this model. The equilibrium values of Q and q increase as k decreases. Unequal mutation rates further increase the equilibrium values and reduce the rates of approach of Q and q to these values. The transitional values of Q and q are very dependent on the initial population frequency composition when mutation rates are unequal. Reducing k , of course, reduces the effective number of alleles, which is further reduced by unequal mutation rates. Complications introduced by initial population composition, unequal mutation rates, and number of allelic states, coupled with data limitations for long-term measures of genetic distance or population differentiation, with mutation as the main driving force, are discussed.

The probability of genes being alike, including both identity by descent and alike in state, is used to formulate the effects of drift within and between monoecious populations in which mutation rates are equal for a finite number of allelic states. The transitional and equilibrium values of the probabilities of alikeness within and between populations are extended to accommodate unequal mutation rates among alleles. Also, the effective number of alleles for this model is compared with that of a stepwise mutation model (1). Some of the results are applied to Nei's (2) measure of genetic distance. Relevant literature, the use of identity by descent for measuring short-term genetic distances, and limitations of data for measuring differentiation among populations are left for the *Discussion*.

Genes Alike Within Populations

Denote the probability of a random pair of genes being alike by Q and unlike by $1 - Q$. Let the general mutation rate be u and the mutation rate of any allele to any other specific allele be v .

A randomly mating monoecious population of constant size N with distinct generations is assumed, so that with probability $1/2N$ a random pair of genes in the offspring stems from a single parental gene and, with probability $1 - (1/2N)$, from different parental genes, but which are alike with probability Q . Whether from a single parental gene or two alike parental genes, the offspring genes are alike with probability $a = (1 - u)^2 + uv$. The term uv is included to allow both genes to mutate to the same allele. If the ancestral genes are not alike, they are alike in the offspring with probability $b = 2(1 - u)v + (u - v)v$ for the cases in which one gene does not mutate and the other mutates to the same gene or both genes mutate to the same gene.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

With these probabilities, a transition equation is constructed (t denotes generation):

$$Q_{t+1} = \frac{a}{2N} + \left(1 - \frac{1}{2N}\right) [Q_t a + (1 - Q_t) b].$$

By equating the Q s, the final or equilibrium value, Q^* , is found,

$$Q^* = \frac{a - b + 2Nb}{2N(1 - \lambda)},$$

where $\lambda = \left(1 - \frac{1}{2N}\right)(a - b)$. The following relationship holds

$$Q^* - Q_t = (Q^* - Q_{t-1})\lambda = (Q^* - Q_0)\lambda^t,$$

so that

$$Q_t = Q^* - (Q^* - Q_0)\lambda^t,$$

where Q_0 is the initial probability of a pair of genes being alike. Note that $Q_0 = 1$ when the initial population is monomorphic.

By dropping terms u^2 and uN^{-1} or less, approximate values are obtained

$$\lambda \cong 1 - 2u - 2v - (1/2N),$$

$$Q^* \cong \frac{1 + 4Nv}{1 + 4Nu + 4Nv}, \quad 1 - Q^* \cong \frac{4Nu}{1 + 4Nu + 4Nv}.$$

With k allelic states and equal probability of mutating from one to the other, $v = u/(k - 1)$ with the result that

$$Q^* = \frac{1 + 4Nu/(k - 1)}{1 + 4Nuk/(k - 1)} = \frac{1 + 4Nv}{1 + 4Nkv}.$$

The probability decreases as k increases and, for the infinite-allele model, the classical result is obtained,

$$Q^* = \frac{1}{1 + 4Nu}.$$

Regarded as a parameter, Q is also the expected frequency of homozygotes in monoecious populations, and $1 - Q$ is the expected frequency of heterozygotes. To estimate Q from a sample of n individuals, there are $n(2n - 1)$ pairs of genes. If we calculate the frequency of pairs alike, we obtain

$$\hat{Q} = \frac{2n \sum_i \hat{p}_i^2 - 1}{2n - 1},$$

where \hat{p}_i is the sample frequency of the i th allele. Further,

$$\mathcal{E}\hat{Q} = Q$$

where \mathcal{E} denotes expectation.

Genes Alike Between Populations

Here we consider two populations from the same founder population that drift independently under the same previous

mutation model. The probability of a random pair of genes, one from each population, being alike is denoted by q and not alike by $1 - q$. Obviously, $q_0 = Q_0$, since the populations were one and the same at time zero.

Using the same arguments as before for the effects of mutation, we find

$$q_{t+1} = q_t a + (1 - q_t) b$$

with an equilibrium value of

$$q^* = \frac{b}{1 - \alpha},$$

where $\alpha = a - b$. Then,

$$q^* - q_t = (q^* - q_{t-1})\alpha = (q^* - q_0)\alpha^t$$

and

$$q_t = q^* - (q^* - q_0)\alpha^t.$$

Dropping the small terms as before gives

$$\alpha \cong 1 - 2u - 2v = 1 - 2uk/(k - 1) = 1 - 2kv$$

$$q^* \cong \frac{v}{u + v} = \frac{1}{k},$$

and q^* is inversely related to the number of allelic states, being zero for the infinite-allele model. The minimum value that q can take is q^* . If the initial population happened to have $q_0 = q^*$, then $q_t = q^*$ is a constant.

If we index the sample allele frequencies of the two populations by 1 and 2, then the frequency with which pairs of alleles are alike is

$$\hat{q} = \sum_i \hat{p}_{i1} \hat{p}_{i2}$$

and

$$\mathcal{E}\hat{q} = q.$$

Unequal Mutation Rates

Unequal mutation rates among alleles modify the results. Let the mutation rate of any other allele to the i th allele be v_i . If $\bar{u} = \sum_i v_i$, the mutation rate of the i th allele to another allele is $\bar{u} - v_i$. Thus, the loss in frequency is $p_i(\bar{u} - v_i)$, while the gain in frequency is $(1 - p_i)v_i$. At equilibrium, there is no change in frequency,

$$p_i(\bar{u} - v_i) = (1 - p_i)v_i \quad \text{or} \quad p_i^* = \frac{v_i}{\bar{u}}.$$

Then, the probability that a random pair of alleles are alike between two independent equilibrium replicate populations is

$$q^* = \sum_i (p_i^*)^2 = \sum_i \frac{v_i^2}{\bar{u}^2} = \frac{k(\bar{v}^2 + \sigma_v^2)}{k^2 \bar{u}^2} = \frac{1 + \frac{\sigma_v^2}{\bar{v}^2}}{k} = \frac{1 + c^2}{k},$$

where $\bar{v} = \bar{u}/k$, σ_v^2 is the variance of the mutation rates, and c is the coefficient of variation of the mutation rates. The effect is to increase q^* or to reduce the effective number of allelic states, which may be taken to be

$$k_e = \frac{k}{1 + c^2}.$$

Unfortunately, we cannot just substitute k_e for k in the previous results.

To develop the transition equation for q , the frequency of

alike pairs of parental alleles i , one from each population, is $\hat{p}_{i1} \hat{p}_{i2}$, and the probability that they remain alike is $a_i = (1 - \bar{u} + v_i)^2 + \sum_{j \neq i} v_j^2$. The frequency of unlike pairs of alleles i and j is $\hat{p}_{i1} \hat{p}_{j2}$, and the probability of their being alike in the offspring is $b_{ij} = (1 - \bar{u} + v_i)v_j + (1 - \bar{u} + v_j)v_i + \sum_{k \neq i, j} v_k^2$. The transition equation becomes

$$q_{t+1} = \mathcal{E} \left(\sum_i \hat{p}_{i1t} \hat{p}_{i2t} a_i + \sum_{i \neq j} \sum \hat{p}_{i1t} \hat{p}_{j2t} b_{ij} \right).$$

Before taking expectation, all possible cancellations are taken after dropping terms \bar{u}^2 or less

$$q_{t+1} = \mathcal{E} \left\{ \sum_i \hat{p}_{i1t} \hat{p}_{i2t} (1 - 2\bar{u}) + \sum_i \hat{p}_{i1t} v_i + \sum_i \hat{p}_{i2t} v_i \right\}.$$

Since $\mathcal{E} \sum_i \hat{p}_{i1t} \hat{p}_{i2t} = q_t$ and $\mathcal{E} \hat{p}_{i1t} = \mathcal{E} \hat{p}_{i2t} = p_{it}$,

$$q_{t+1} = q_t (1 - 2\bar{u}) + 2 \sum_i p_{it} v_i$$

and the transition is gene-frequency-composition dependent,

$$q^* - q_{t+1} = (q^* - q_t)\alpha + 2\bar{v}(1 + c^2) - 2 \sum_i p_{it} v_i,$$

where $\alpha = 1 - 2\bar{u}$.

Denote the mean of the v_i s as

$$\bar{v}_t = \sum_i p_{it} v_i.$$

Then, $\bar{v}(1 + c^2) = \bar{v}^*$ is the equilibrium value. To find the transition equation for \bar{v} , note that

$$p_{i,t+1} = p_{it}(1 - \bar{u} + v_i) + (1 - p_{it})v_i$$

and

$$\sum_i p_{i,t+1} v_i = \sum_i p_{it} v_i (1 - \bar{u}) + \sum_i v_i^2$$

or

$$\bar{v}_{t+1} = \bar{v}_t (1 - \bar{u}) + k\bar{v}_t^2 (1 + c^2).$$

Using $k\bar{v} = \bar{u}$ and $\bar{v}(1 + c^2) = \bar{v}^*$, we find

$$\bar{v}^* - \bar{v}_{t+1} = (\bar{v}^* - \bar{v}_t)\rho = (\bar{v}^* - \bar{v}_0)\rho^{t+1},$$

where $\rho = 1 - \bar{u}$. If the initial population is fixed for the i th allele, then $\bar{v}_0 = v_i$.

We now substitute these results into the transition equation for q ,

$$q^* - q_{t+1} = (q^* - q_t)\alpha + 2(\bar{v}^* - \bar{v}_0)\rho^t.$$

Repeated substitution into $q^* - q_t$ produces

$$q^* - q_{t+1} = (q^* - q_0)\alpha^{t+1} + 2(\bar{v}^* - \bar{v}_0) \sum_{i=0}^t \alpha^i \rho^{t-i}.$$

After evaluating

$$\sum_{i=0}^t \alpha^i \rho^{t-i} = \frac{\rho^{t+1} - \alpha^{t+1}}{\rho - \alpha}, \quad \rho - \alpha = \bar{u},$$

the transitional value of q is found

$$q_t = q^* - (q^* - q_0)\alpha^t - 2(\bar{v}^* - \bar{v}_0)(\rho^t - \alpha^t)/\bar{u}.$$

Again, if $q_0 = q^*$, then $\bar{v}_0 = \bar{v}^*$ and $q_t = q^*$ remains constant.

We now sketch the transition for alikeness within populations. Before mutation, the offspring genes that stem from a single parental gene are alike for the i th allele with frequency

\hat{p}_i and those from two distinct parental genes are alike for the i th allele with frequency $(2N\hat{p}_i^2 - \hat{p}_i)/(2N - 1)$. The frequency of unlike pairs of parental alleles i and j is $2N\hat{p}_i\hat{p}_j/(2N - 1)$. The transition equation becomes

$$Q_{t+1} = \mathcal{E} \left\{ \frac{\sum_i \hat{p}_{it} a_i}{2N} + \left(1 - \frac{1}{2N} \right) \left[\sum_i \frac{(2N\hat{p}_{it}^2 - \hat{p}_{it})}{2N - 1} a_i + \frac{2N}{2N - 1} \sum_{i \neq j} \hat{p}_{it} \hat{p}_{jt} b_{ij} \right] \right\}$$

Again, all possible cancellations are taken after dropping terms \bar{u}^2 and $\bar{u}N^{-1}$ or less.

$$Q_{t+1} = \mathcal{E} \left\{ \frac{1 + 4N \sum_i \hat{p}_{it} v_i}{2N} + \left(1 - \frac{1}{2N} \right) \times \frac{\left(2N \sum_i \hat{p}_{it}^2 - 1 \right)}{2N - 1} \left(1 - 2\bar{u} \right) \right\}$$

Since

$$\mathcal{E} \frac{2N \sum_i \hat{p}_{it}^2 - 1}{2N - 1} = Q_t,$$

we have

$$Q_{t+1} = Q_t \lambda + \frac{1 + 4N \sum_i p_{it} v_i}{2N} = Q_t \lambda + \frac{1 + 4N \bar{v}_t}{2N}$$

and at equilibrium

$$\bar{v}^* = \bar{v}(1 + c^2), \quad Q^* \cong \frac{1 + 4N\bar{v}(1 + c^2)}{1 + 4Nk\bar{v}},$$

which reduces to the previous Q^* when $c^2 = 0$ and $\bar{v} = v$. Again, the equilibrium value is increased with unequal mutation rates as expected.

The transition equation is also gene-frequency-composition dependent,

$$(Q^* - Q_{t+1}) = (Q^* - Q_t)\lambda + 2(\bar{v}^* - \bar{v}_t).$$

Substituting λ for α and Q for q in the derivation of q , we get

$$Q_t = Q^* - (Q^* - Q_0)\lambda^t - (\bar{v}^* - \bar{v}_0)2(\rho^t - \lambda^t)/(\rho - \lambda),$$

$$\rho - \lambda = (1 + 2Nk\bar{v})/2N.$$

While the effects of unequal mutation rates on equilibrium values q^* and Q^* are readily apparent, the effects on transitional values are not and require some numerical evaluations for clarification.

As an example, five allelic states are considered with mutation rates of 10^{-3} , $10^{-3.5}$, 10^{-4} , $10^{-4.5}$, and 10^{-5} . The average mutation rate is $\bar{v} \cong 2.92 \times 10^{-4}$ and $1 + c^2 \cong 2.614$, $q^* \cong 0.523$. Note that q^* and c^2 do not change for any constant, say 10^{-2} , times the mutation rates. For Q , two population

sizes, 100 and 1,000, are considered, which have the following characteristics

N	$4N\bar{u}$	Q^*
100	0.58	0.824
1,000	5.83	0.593

Two initial frequencies are used, $q_0 = Q_0 = 0.2, 1$. The 0.2 is for a case when the initial values are less than the equilibrium values and implies equal gene frequencies with $\bar{v}_0 = \bar{v}$. For $q_0 = 1$, initial populations are considered fixed for the allele with the largest mutation rate to other alleles, $\bar{v}_0 = 10^{-5}$, and for the allele with the smallest mutation rate to other alleles, $\bar{v}_0 = 10^{-3}$. Also considered for $q_0 = 1$ is the equal mutation model with $v = \bar{v}$. For this model, $q^* = 0.2$, $Q^* = 0.705$ ($N = 100$), 0.317 ($N = 1,000$).

The results for q are displayed in Fig. 1 and for Q in Fig. 2. With equal mutation rates, the transitional rate of approach of q to q^* is constant, α . With unequal mutation rates, the transitional rate is not constant. For $\bar{v}_0 > \bar{v}^*$, as in the case of $\bar{v}_0 = 10^{-3}$, the transitional rate is slower throughout as expected, although the difference from equal mutation rates is not so obvious because of the difference in equilibrium values. With $\bar{v}_0 < \bar{v}^*$, as in the case of $\bar{v}_0 = 10^{-5}$, the initial transitional rate is faster than for equal mutation rates. However, the equilibrium value is overshoot and the return toward q^* from below is at a slower rate than for equal mutation rates. Also, for $\bar{v}_0 = \bar{v} < \bar{v}^*$, the approach to q^* is from below and at a slower rate than for equal mutation rates. Thus, the general effect, particularly in the long run, of variable mutation rates is to reduce the transitional rate.

The same types of effects of unequal mutation rates on the transition of Q are seen in Fig. 2. The equilibrium values are of course approached at a much more rapid rate for $N = 100$ (Fig. 2a) than for $N = 1,000$ (Fig. 2b).

Effective Number of Alleles

The so-called effective number of alleles, m , at equilibrium is given by $m = 1/Q^*$ or

$$m_k = \frac{1 + 4Nuk/(k - 1)}{1 + 4Nu/(k - 1)} = \frac{1 + 4Nkv}{1 + 4Nv}$$

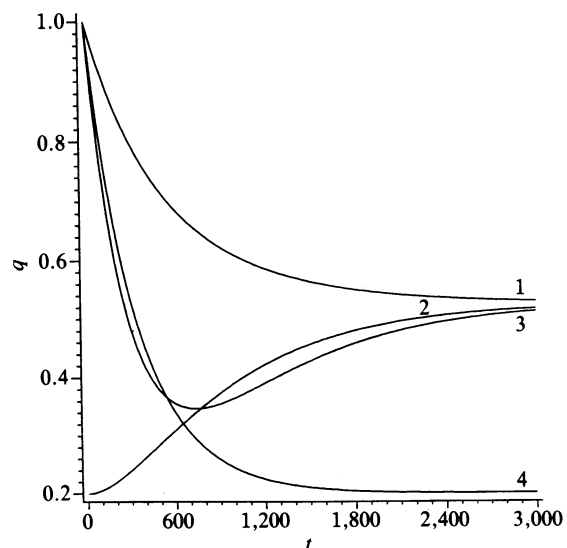


FIG. 1. Transitional values of q for various initial conditions \bar{v}_0 and q_0 . Curves: 1, $\bar{v}_0 = 10^{-3}$, $q_0 = 1$; 2, $\bar{v}_0 = \bar{v}$, $q_0 = 0.2$; 3, $\bar{v}_0 = 10^{-5}$, $q_0 = 1$; 4, equal mutation rates, $q_0 = 1$.

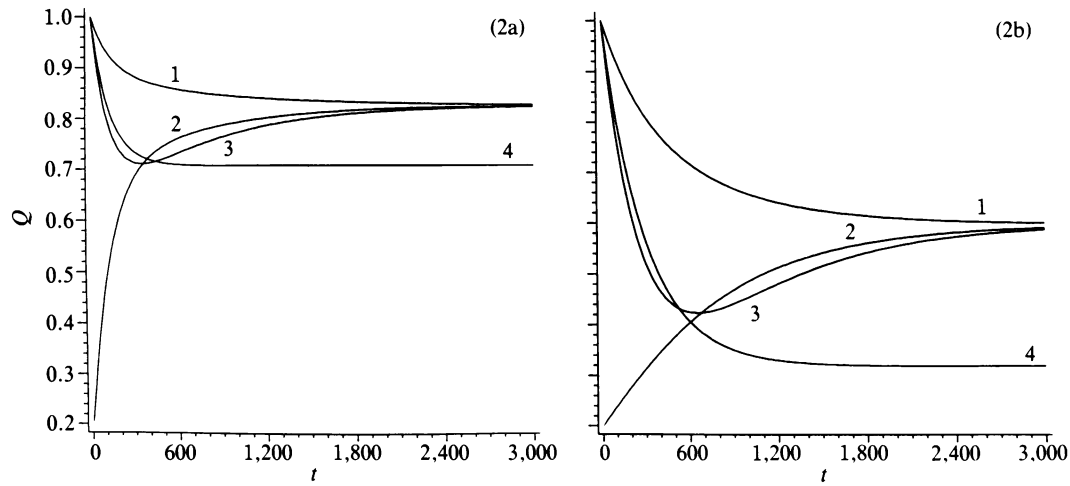


FIG. 2. Transitional values of Q for various initial conditions \bar{v}_0 and Q_0 . (a) $N = 100$. (b) $N = 1,000$. Curves: 1, $\bar{v}_0 = 10^{-3}$, $Q_0 = 1$; 2, $\bar{v}_0 = \bar{v}$, $Q_0 = 0.2$; 3, $\bar{v}_0 = 10^{-5}$, $Q_0 = 1$; 4, equal mutation rates, $Q_0 = 1$.

for the k -allelic-state model, which reduces to the classical result for the infinite-allele model

$$m_\infty = 1 + 4Nu$$

(3). With unequal mutation rates, the effective number of alleles is reduced to

$$m_k = \frac{1 + 4Nk\bar{v}}{1 + 4N\bar{v}(1 + c^2)}$$

Ohta and Kimura (1), in studying a specific stepwise mutation model for electrophoretic variants, found the effective number of alleles to be

$$m_s = \sqrt{1 + 8Nu}$$

By equating m_k to m_s , an equivalent number, k' , of allelic states is found to be a solution to $(k - 1)(k - 3) = 8Nu$ or

$$k' = 2 + \sqrt{1 + 8Nu}$$

Thus $k' > 3$ and can be much larger when $4Nu$ is very large but is only 4 when $4Nu = 1.5$, 5 when $4Nu = 4$, and so on. Ohta and Kimura (1) noted that m_s and m_∞ are not very different for small $4Nu$ but that m_s is much smaller than m_∞ for large $4Nu$.

Nei's Distance Measure

Nei (2) defined a distance measure, $D = -\ln I$, where after correcting for bias (4) and reducing the formula to that for a single locus

$$I = \frac{\hat{q}}{\sqrt{\hat{Q}_1\hat{Q}_2}}$$

The expectation of I is approximated in the following manner:

$$\begin{aligned} \mathcal{E}I &\cong \frac{\mathcal{E}\hat{q}}{\sqrt{\mathcal{E}\hat{Q}_1\mathcal{E}\hat{Q}_2}} = \frac{q}{Q} \\ &= \frac{q^* - (q^* - Q_0)\alpha^t - (\bar{v}^* - \bar{v}_0)2(\rho^t - \alpha^t)/(\rho - \alpha)}{Q^* - (Q^* - Q_0)\lambda^t - (\bar{v}^* - \bar{v}_0)2(\rho^t - \lambda^t)/(\rho - \lambda)}. \quad [1] \end{aligned}$$

There is considerable simplification if the founder population is in equilibrium with respect to mutation and drift—i.e., $Q_0 = Q^*$ and $\bar{v}_0 = \bar{v}^*$. Then,

$$\mathcal{E}I = \frac{q^* - (q^* - Q^*)\alpha^t}{Q^*},$$

but the relationship to time is still complex with a few allelic states. Returning to the original formula (1), there is simplification for an infinite number of allelic states,

$$\mathcal{E}I = \frac{Q_0\alpha^t}{Q^* - (Q^* - Q_0)\lambda^t},$$

but again the relationship to time is complex if the original population is not in equilibrium. It is only with the infinite-allele model and initial equilibrium that the relationship to time simplifies.

$$\mathcal{E}I = \alpha^t, \quad -\ln \mathcal{E}I \cong 2ut.$$

Discussion

Many of these results are not new, particularly for equal mutation rates. For the infinite-allele model, Q^* was given by Kimura and Crow (3) and solutions for both Q and q were given by Nei (5). With k alleles, Q^* and q^* were obtained by Kimura (6). Also, Q^* can be found in Watterson (7) and is a special case of Takahata's (8) study of a composite stepwise mutation model. The transition equation for Q is given by Golding and Strobeck (9) as $\Phi_{1/0}$, corresponding to a single site with k alleles, in their study of the distribution of nucleotide site differences. Also, from general multiple-site formulations of Takahata (10), by focusing on a single site, the equilibrium value of Q and the transition equation for q can be deduced. In this context, q and Q are for a single variant (site) with k alleles (states).

Griffiths (11) developed a solution for Q with equal mutation rates between pairs of alleles but different among pairs. The effects of unequal mutation rates are not readily apparent in his formulations. By an inequality, he shows for a constant overall mutation rate that Q is least when all mutation rates are equal. Our model and results are the same as his with equal mutation rates. Griffiths (12) also formulates q for his mutation model but with the initial population in equilibrium ($q_0 = Q^*$). Again, he shows by inequalities that q is least for equal mutation rates.

The effects of unequal mutation rates on q^* and Q^* are readily apparent for our model, increasing both. Unequal mutation rates reduce the transitional rates of q and Q , at least in the long run. The numerical illustrations of the transitions demonstrate the key role that the initial population composition plays, leading to a variety of transitional outcomes.

The frequencies q and Q are dependent on the measuring device whether it be of electrophoretic variants or other means of identifying allelic variation. The appropriate mutation rates are those that change one recognizable variant into another recognizable variant. In the case of electrophoretic variants, it does not matter whether the changes are conformational, electrical charge, or other, as long as they are recognized and are mendelian. In the same vein, all those changes, even those known to occur, that are not recognized by the measuring device are appropriately ignored. Thus, two different electrophoretic protocols for the same enzyme will often have different parameters. These factors will lead to differences among loci in addition to other inherent differences among loci such as mutation rates and numbers of allelic states.

The greater the number of allelic states, the more information there is on the differentiation of populations and species. Singh *et al.* (13), by varying gel concentration and pH and with the use of heat denaturation, increased the number of allelic classes from 6 by standard gel electrophoresis to 37 for xanthine dehydrogenase in *Drosophila pseudoobscura* and turned up unsuspected population differences. In most studies, however, standard protocols are used. Even when a large net is cast in terms of the geographic distribution of populations or species (or both), the numbers of allelic states are often small. Whether real or from limitations of the measuring devices, the limitations of the data are real in terms of the number of allelic states for genetic distances or other measures of population differentiation.

An alternative measure of distance may be based on identity by descent for short-term evolution. It is often not clear to me when reading the literature whether "identity by descent," "alike in state," or some combination is being invoked when the term "identity" is used. Most often identity appears to mean simply alike. Without mutation, identity by descent is the most useful concept for describing the effects of drift. For example, without mutation,

$$Q_t = Q^* - (Q^* - Q_0)\lambda^t$$

as before, except now $Q^* = 1$ and $\lambda = 1 - 1/2N$, so that

$$Q_t = 1 - \lambda^t + \lambda^t Q_0 = \theta_t + (1 - \theta_t)Q_0,$$

where $\theta_t = 1 - \lambda^t$ is the probability that a random pair of genes are identical by descent. For this model, $q_t = q_0 = Q_0$. Consequently, an estimate of θ is provided by

$$\hat{\theta} = \frac{\hat{Q}_1 + \hat{Q}_2 - 2\hat{q}}{2(1 - \hat{q})}$$

for equal-sized samples from the two populations, regardless of Q_0 and the measuring device as long as there are at least two allelic states. It is unbiased in the sense that the ratio of expectations is unbiased. While not readily apparent, this is the same estimate of θ from pooling the within-individual and between-individual-within-population mean squares, as is appropriate for monoecious populations in which the inbreeding and coancestry coefficients are the same (14) and further summing numerators and denominators over alleles. If one omits the sample size corrections in θ , it reduces to Latter's (15) estimate of kinship, φ^* . It is this estimator com-

bined over loci that was used as a basis for a distance measure

$$\hat{\mathcal{D}} = -\ln(1 - \hat{\theta}), \quad \mathcal{E}\hat{\mathcal{D}} \cong t/2N$$

proposed and investigated by Reynolds *et al.* (16) as a measure of short-term evolution before mutations have an over-riding effect.

The number of recognizable allelic states affects the estimate of θ in terms of bias and sampling variance. Bias decreases as the number of states increases but can essentially be eliminated by summing \hat{Q} s and \hat{q} s over several loci. That having been accomplished, the information about θ becomes roughly aligned with the total degrees of freedom, $\sum_i (k_i - 1)$ (i indexes loci), for states (16). Mutational synonymities that are not recognized, as is often the case with electrophoresis, actually extend the time scale over which identity by descent is an appropriate measure.

Long-term genetic distance measures with mutation rates as the main driving force are much more complicated. If one knows k and q_0 , then

$$-\frac{(k-1)}{k} \ln \left(\frac{q - \frac{1}{k}}{q_0 - \frac{1}{k}} \right) = 2ut$$

is linearly related to divergence time as a multiple of the mutation rate [considered by Takahata (10) in a more complex setting]. Without making some assumption about q_0 , the task is impossible. Even if one assumes equilibrium values, $q_0 = Q^*$, the linear relationship of Nei's D to divergence time depends on a large number of allelic states, which in practice is determined by the measuring device. Nei's distance measure is considered only as an example. It is anticipated that other measures of genetic distance or of population differentiation based on gene frequencies and mutations will have even greater difficulties.

In applications, \hat{q} s and \hat{Q} s are summed or averaged over loci for the estimation of genetic distance, many loci being required to reduce the sampling variability to acceptable terms. An average for many loci does not alleviate the problems concerning initial population composition, unequal mutation rates, and the number of allelic states, however.

I thank Drs. W. J. Ewens, J. Felsenstein, G. B. Golding, N. Takahata, and B. S. Weir for considerable help in the development of this manuscript. This investigation was supported in part by Research Grant GM 11546 from the National Institute of General Medical Sciences. This is paper no. 8888 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, N.C.

- Ohta, T. & Kimura, M. (1973) *Genet. Res. (Cambridge)* **22**, 201-204.
- Nei, M. (1972) *Am. Nat.* **106**, 283-292.
- Kimura, M. & Crow, J. F. (1964) *Genetics* **49**, 725-738.
- Nei, M. (1978) *Jpn. J. Human Genet.* **23**, 341-369.
- Nei, M. (1976) in *Population Genetics and Ecology*, eds. Karlin, S. & Nevo, E. (Academic, New York), pp. 723-765.
- Kimura, M. (1968) *Genet. Res. Camb.* **11**, 247-269.
- Watterson, G. A. (1977) *Genetics* **85**, 789-814.
- Takahata, N. (1980) *J. Mol. Evol.* **15**, 13-20.
- Golding, G. B. & Strobeck, C. (1982) *Theor. Popul. Biol.* **22**, 96-107.
- Takahata, N. (1982) *Genet. Res. Camb.* **39**, 63-77.
- Griffiths, R. C. (1980) *Theor. Popul. Biol.* **17**, 51-70.
- Griffiths, R. C. (1980) *J. Math. Biol.* **10**, 195-204.
- Singh, R. S., Lewontin, R. C. & Felton, A. A. (1976) *Genetics* **84**, 609-629.
- Cockerham, C. C. (1969) *Evolution* **23**, 72-84.
- Latter, B. D. H. (1973) in *Population Genetics Monographs, Volume III. Genetic Structure of Populations*, ed. Morton, N. E. (Univ. Press of Hawaii, Honolulu), pp. 27-29.
- Reynolds, J., Weir, B. S. & Cockerham, C. C. (1983) *Genetics* **105**, 767-779.